贝叶斯方差分析在 JASP 中的实现

王允宏 [1] Don van den Bergh[2] Frederik Aust[2] Alexander Ly[3] Eric-Jan Wagenmakers[2] 胡传鹏 [4*]

([1] 上海师范大学心理学系，上海 200234；[2]Department of Psychological Methods, University of Amsterdam, 1018 VZ Amsterdam, The Netherlands；[3]Centrum Wiskunde & Informatica, 1090 GB Amsterdam, The Netherlands；[4] 南京师范大学心理学院，南京 210024)

摘 要：贝叶斯统计应用于假设检验的方法——贝叶斯因子——在心理学研究中的应用日渐增加。贝叶斯因子能分别量化所支持的相应假设或模型的证据，进而根据其数值大小做出当前数据更支持哪种假设或模型的判断。然而，国内尚缺乏对方差分析的贝叶斯因子的原理与应用的介绍。基于此，本文首先介绍贝叶斯方差分析的基本思路及计算原理，并结合实例数据，展示如何在 JASP 中对五种常用的心理学实验设计(单因素组间设计、单因素组内设计、二因素组间设计、二因素组内设计和二因素混合设计)进行贝叶斯方差分析及如何汇报和解读结果。贝叶斯方差分析提供了一个能有效替代传统方差分析的方案，是研究者进行统计推断的有力工具。

关键词：贝叶斯统计；贝叶斯因子；方差分析；JASP

1 引言

方差分析(ANOVA)这一统计方法适用于评估分类型预测变量(自变量)对连续型结果变量(因变量)的影响，是实验心理学中常用的统计方法(Fritz et al., 2012)。在零假设显著性检验(null hypothesis significance testing, NHST)的框架下，方差分析得到的结果会根据 p 值进行统计显著性推断：当设定了 p 值阈限后，研究者往往会根据 p 值，以全或无的方式推断结

* 通讯作者：胡传鹏，男，博士，南京师范大学心理学院教授，E-mail: hcp4715@hotmail.com

果的统计显著性。例如，如果 $p < 0.05$，就说明结果具有统计显著性，应当拒绝零假设(H_0)；如果 $p > 0.05$，就说明结果不具有统计显著性。这种二分的观点受到了广泛质疑，并且这也是心理学可重复性危机的来源之一(Open Science Collaboration, 2015; Schmalz et al., 2021)。因此，贝叶斯统计(Bayesian statistics)作为一种替代 NHST 的方法，逐渐受到了研究者的关注(Wagenmakers et al., 2011)。

贝叶斯统计的基本思想是随着观测数据的积累，信念(知识经验)不断更新的过程(Faulkenberry et al., 2020; van den Bergh et al., 2020; Wagenmakers, Marsman, et al., 2018)。研究者在进行假设检验前可能会有多个相互竞争的假设，信念即是对各个假设为真的可能性的估计。当某个假设能很好的预测数据时，与该假设一致的信念会得到增强；反之，当某假设对观测数据的预测性很差时，信念就会减弱。因此通过贝叶斯统计，研究者可以分别得到支持 H_1 和 H_0 的证据，进而量化两种假设相对出现的可能性，即通过模型比较(model comparison)的方式来得到贝叶斯因子(Bayes factors, BF; 胡传鹏 等, 2018; 李贵玉, 顾昕, 2021; 许岳培 等, 2022)。虽然贝叶斯统计具有量化对 H_1 和 H_0 的支持程度、不依赖于抽样计划等优势(Grünwald et al., 2020; Hendriksen et al., 2021; Schmalz et al., 2021; Wagenmakers, Marsman, et al., 2018)，但相比于频率主义 ANOVA，贝叶斯 ANOVA 的应用是有限的。随着具有图形界面的软件(例如，JASP)的开发，BF 的使用变得更加简便，因此也开始被广泛应用于心理学的各个领域(孟迎芳 等, 2021; Brydges & Gaeta, 2019; Derks et al., 2021; Rouder et al., 2017; Wagenmakers, Love, et al., 2018)。先前亦有中文文献对 JASP 进行了相关介绍，例如，胡传鹏等(2018)文章中 3.1 部分。如果需要了解更多关于 NHST 与 BF 的内容，可参考前人研究(Kruschke & Liddell, 2018; Schmalz et al., 2021; Tendeiro & Kiers, 2019)。

然而，在先前关于 BF 的中文介绍中，多以相关分析和 t 检验作为例子(胡传鹏 等, 2018; 吴凡 等, 2018)。虽然容易理解，但无法应用于方差分析的情境中。主要原因之一在于，贝

叶斯 ANOVA 以贝叶斯的线性模型(linear model, LM)为基础(Liang et al., 2008; Rouder et al., 2012)。虽然也可以使用贝叶斯广义线性模型(Generalized linear model, GLM)，但当前版本的 JASP 并不能进行贝叶斯 GLM 的相关计算。贝叶斯 ANOVA 通过多个不同的线性模型相互之间的比较获得 BF 值。不同模型可能涉及到不同的模型构建方式，这就导致了相比于贝叶斯 t 检验，模型比较的过程更为复杂。此外，随着自变量的增加，模型比较及之后产生的模型选择不确定性的问题会对研究者汇报和解释结果造成困扰。因此，如何解决这种情况下的 BF 计算与解读，也需要额外的知识。为解决研究者在使用贝叶斯 ANOVA 时可能出现的上述问题，本文将介绍贝叶斯线性模型及模型比较的基础知识，并介绍贝叶斯模型平均法(Bayesian model averaging, BMA)这一解决自变量较多时如何计算贝叶斯因子的方法。为了方便没有代码基础的研究者执行数据分析，本文使用了 JASP 这款开源、免费和具有图形界面的统计软件(JASP team, 2022)。如果读者仅关注如何使用 JASP 来进行贝叶斯 ANOVA 分析，以及如何解释和汇报输出的结果，可跳过本文的 2.2 小节。

2 基本概念

2.1 贝叶斯因子

BF 是一种模型比较和选择的方法，反映了对某一模型支持程度的量化(Schmalz et al., 2021)。它等价于观测数据中两个模型的边际似然之比(marginal likelihood ratios; Heck et al., 2022)。在应用于假设检验中时，假设之间的比较可视作模型之间的比较。因此，BF 可用来衡量对 H_1 和 H_0 的支持程度。具体而言，将 H_0 指定为零模型(null model)M_0 并将 H_1 指定为备择模型(alternative model)M_1。BF_{10} 表示相对于 M_0，观测数据对 M_1 的支持程度。例如，$BF_{10} = 12$ 表示观测数据支持 H_1 为真的程度是支持 H_0 为真的程度的 12 倍。如表 1 所示，参考先前研究者对 BF 数值大小所代表意义的划分(胡传鹏 等, 2018; Jarosz & Wiley, 2014;

Jeffreys, 1961; Wetzels et al., 2011)，$BF_{10} = 12$ 可被解释为观测数据提供了较强的证据支持 H_1 为真。反之，$BF_{01} = 12$ 可被解释为观测数据提供了较强的证据支持 H_0 为真。

表 1 BF_{10} 数值划分及其代表意义

BF_{10}	**代表意义**
> 100	极强的证据支持 H_1
30 – 100	非常强的证据支持 H_1
10 – 30	较强的证据支持 H_1
3 – 10	中等程度的证据支持 H_1
1 – 3	较弱的证据支持 H_1
1	无证据
1/3 – 1	较弱的证据支持 H_0
1/10 – 1/3	中等程度的证据支持 H_0
1/30 – 1/10	较强的证据支持 H_0
1/100 – 1/30	非常强的证据支持 H_0
< 1/100	极强的证据支持 H_0

BF 计算的一般公式为：

$$BF_{10} = \frac{p(data|M_1)}{p(data|M_0)} \tag{1)}$$

其中，$p(data|M_1)$表示边际似然，即当前数据在模型M_1中出现的可能性，$p(data|M_0)$同理。因此，BF_{10}反映了两个模型的边际似然之比。关于上式的进一步展开形式，见在线补充材料 A(在线补充材料详见 https://osf.io/7caju/)。

2.2 贝叶斯 ANOVA 中线性模型的比较

要理解 BF 在 ANOVA 中的应用，首先需要理解方差分析与线性模型的关系，原因在于贝叶斯 ANOVA 中 BF 值的计算是基于不同线性模型之间的比较。作为线性模型的一种特殊形式，ANOVA 涉及的自变量是分类变量，因变量是连续变量，且误差项(error)需要满足正态分布。它的特殊之处在于：由于自变量为分类变量(例如，性别、不同实验条件等)，其设计矩阵(design matrix)中包含的元素均为 0 或 1。线性模型的一般形式为：

$$Y_{ij} = \mu + \beta_1 X_{i1} + \ldots + \beta_j X_{ij} + \epsilon_{ij} \ \ (i = 1,2,\ldots,n) \tag{2}$$

其中，Y_{ij}表示因变量，即 j 组个体 i 的数据；X_{ij}表示自变量，取值为 0 或 1，代表个体 i 属于组 j；β_j表示自变量的效应，即某个实验处理的效应；μ表示截距项；ϵ_{ij}表示随机误差，即因变量无法被自变量解释的部分。

假设存在一个两水平的自变量 A，且为了使参数不受计量单位的影响，需要将参数进行标准化$\theta = \frac{\beta}{\sigma}$($\sigma$代表标准误)，即转化为效应量(effect size)。

那么计算 A 存在效应的线性模型可写成：

$$H_1: Y = \mu + \sigma\theta X + \epsilon \tag{3}$$

不包括 A 效应的零模型写作：

$$H_0: Y = \mu + \epsilon \tag{4}$$

如果研究者关注 A 的主效应，在传统的 ANOVA 中，可以通过方差分解的方式来计算 F 值和 p 值，再在零假设显著性检验框架下进行统计推断。在 BF 分析中，研究者是计算当前数据出现在 H_0 和 H_1 这两个模型下的可能性的比例。也就是说，研究者将认为 A 主效应存在的 H_1 指定为 M_1，认为 A 主效应不存在的 H_0 指定为 M_0，通过模型比较的方式来计算出数据支持两个模型可能性的比值，得到 BF_{10} 或 BF_{01}。

上述内容描述的是仅有一个自变量的情况。当存在两个自变量 X_a 和 X_b 时，潜在线性模型的数量增加。从完全不包括任何自变量效应的模型(零模型)到包括全部自变量效应的模型(两个自变量的主效应及其交互作用，即全模型)，共包括 5 个模型(模型的构建请见补充材料 B)。如果以零模型作为 H_0 的模型，则与传统的 ANOVA 仅有两主效应与一个交互作用的 F 值(及 p 值)不同，BF 分析中会报告 4 个 BF 值。这是因为仅包括两个主效应但无交互作用的模型(见线上补充材料 B 中的公式 12)也会与零模型进行比较从而得到 BF 值。并且，包括交互作用的模型也同时包括两个主效应，而非单纯地只包括交互作用。此外如果研究

者想计算交互作用所对应的 BF 值，可以应用 BF 的传递性(Srinivasan & Vijayaragunathan, 2021; Wagenmakers et al., 2010)。具体计算原理请见补充材料 C。

由此可见，随着自变量个数的增加，模型的数量也会迅速增加。这就导致研究者很难单独地考虑每个模型的效应。同时，模型两两比较产生的 BF 数量也随之增加，这会对研究者解释结果造成困扰。当H_1选择不同的模型时，与H_0对应的原模型比较产生可能会产生几个数值相似的BF_{10}时(即有相同程度的证据支持几个模型作为H_1)，研究者该如何选择合适的模型作为H_1进而得出结论？当实验中存在 2 个自变量时，可供H_1选择的备择模型有 4 个；当存在 3 个自变量时，可供选择的备择模型就增长到了 18 个。由此可见，传统 ANOVA 得出的 F 值和 p 值的数量少于 BF 的数量。这就造成了研究者在模型选择上会出现困难。研究者甚至可能会为了追求 BF 值最大化，选取特定的模型，从而忽略了模型选择的不确定性。据此，研究者提出使用贝叶斯模型平均(Bayesian model averaging, BMA)的方法来解决这一问题(Heck & Bockting, 2021; Heck et al., 2022; Hinne et al., 2020; van den Bergh et al., 2020; Wagenmakers, Love, et al., 2018)。关于 BMA 的详细信息请见在线补充材料 D。

2.3 JASP 中进行贝叶斯 ANOVA

本文使用的 JASP 版本为 0.16.3(JASP Team, 2022)，所用示例数据及分析结果可在 OSF 上获取(https://osf.io/7caju/)。JASP 是一个不断开发中的软件，开发者不断地改进该软件并且修复现有版本中的问题。因此，我们推荐读者使用最新版的 JASP 而非一定要使用本版本。JASP 使用 R 中的 BayesFactor 包(Morey & Rouder, 2022)计算 BF。在进行贝叶斯 ANOVA 前，研究者需要进行的基本步骤是：指定自变量和因变量、选择输出的 BF 形式。如果是多自变量的实验设计，还需设置BF_{incl}的输出形式。

不同于传统 ANOVA 的输出指标(F, p, η^2等)，在 JASP 中进行贝叶斯 ANOVA 后，会输

出如下指标(见图 2)：Model Comparison 为模型比较的结果汇总表；$P(M)$表示获得观测数据前模型的先验概率，在默认的 Uniform 分布中每个模型的先验概率相同，即如果有 m 个模型，每个模型的先验概率就为 1/m；$P(M|data)$表示获得数据后模型的后验概率；BF_M表示当前模型从先验到后验的变化：

$$BF_M = \frac{P(M|data)}{1 - P(M|data)} \times \frac{1 - P(M)}{P(M)} \quad 5)$$

Error %表示结果的变异系数。由于在计算过程中使用了 MCMC 采样，因此每重新运行一次分析，结果都可能会出现细微的差异。因此，按照本文步骤，读者所做的贝叶斯 ANOVA 结果可能会与本文得到的结果有细微不同。结果的波动可用 Error %来量化，Error %越高说明结果的波动性越高。van Doorn 等(2021)推荐 Error %小于 20%时通常是可以接受的。

3 使用 JASP 进行贝叶斯 ANOVA 实例展示

3.1 单因素方差分析

示例数据。使用疼痛阈限(Pain Thresholds)数据作为示例，该数据来自单因素组间设计的实验。自变量为发色(Hair Color)，包括 4 个水平(见图 1)：深色黑发(Dark Brunette)、深色金发(Dark Blond)、浅色金发(Light Blond)、浅色黑发(Light Brunette)。因变量为疼痛容忍度(Pain Tolerance)。因此在本例中，H_0 为发色对疼痛阈限无影响，H_1 为发色对疼痛阈限有影响。例如，浅色金发个体对疼痛的容忍度要高于浅色黑发个体。



图 1 JASP 中示例数据疼痛阈限的描述统计结果

操作步骤。首先在 JASP 中打开数据(Open - Data Library - ANOVA - Pain Thresholds)，然后在 ANOVA 面板中选择“Bayesian AVOVA”。 之后需要进行以下步骤(见图 2)：第一，将疼痛容忍度作为因变量放入“Dependent Variable”中，将发色作为自变量放入“Fixed Factors”中；第二，在“Bayes Factor”选项框中选择要输出的 BF 形式；第三，在“order”选项框中选择模型比较的顺序，即确定模型是与零模型(null model)还是与最优模型(best model)进行比较。如果选择“compare to best model”，那么“Model Comparison”中呈现结果的第一行就代表最优模型与其自身比较的结果。反之，如果选择“compare to null model”，那么就代表零模型与其自身比较的结果。因此“order”的选择对结果没有实际影响；第四，由于该示例数据中的自变量发色是一个四水平组间变量，因此需要事后检验(Post Hoc Tests)来观测差异究竟出现在哪两个水平之间。然而，事后检验本身是一个探索性的分析过程。因此，本文建议研究者在进行数据分析前，需要做出明确假设。

贝叶斯方差分析中的事后检验无需进行校正(Gelman et al., 2012)。不同于依赖预设显著

性水平以及 p 值的 NHST，贝叶斯方差分析关注参数或模型的后验分布及观测数据对不同假设的支持程度。贝叶斯方差分析提供了完整的后验分布，使得研究者可以直接对不同模型的后验分布进行比较。并且，贝叶斯方差分析允许研究者设定不同的先验分布形式，来反映对不同假设的先验信念。贝叶斯因子的本质是两个模型边际似然的比值，量化了对模型的相对支持程度，直观的对模型进行了比较。综上，在贝叶斯方差分析中无需使用多重比较校正。

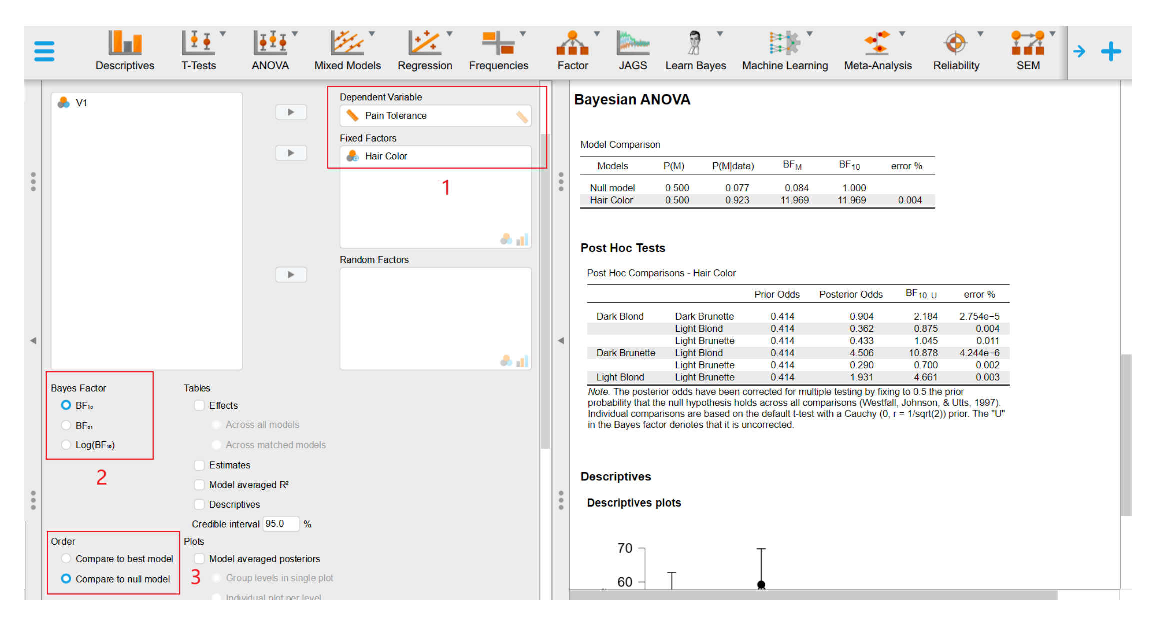


图 2 使用 JASP 进行单因素方差分析的操作截屏

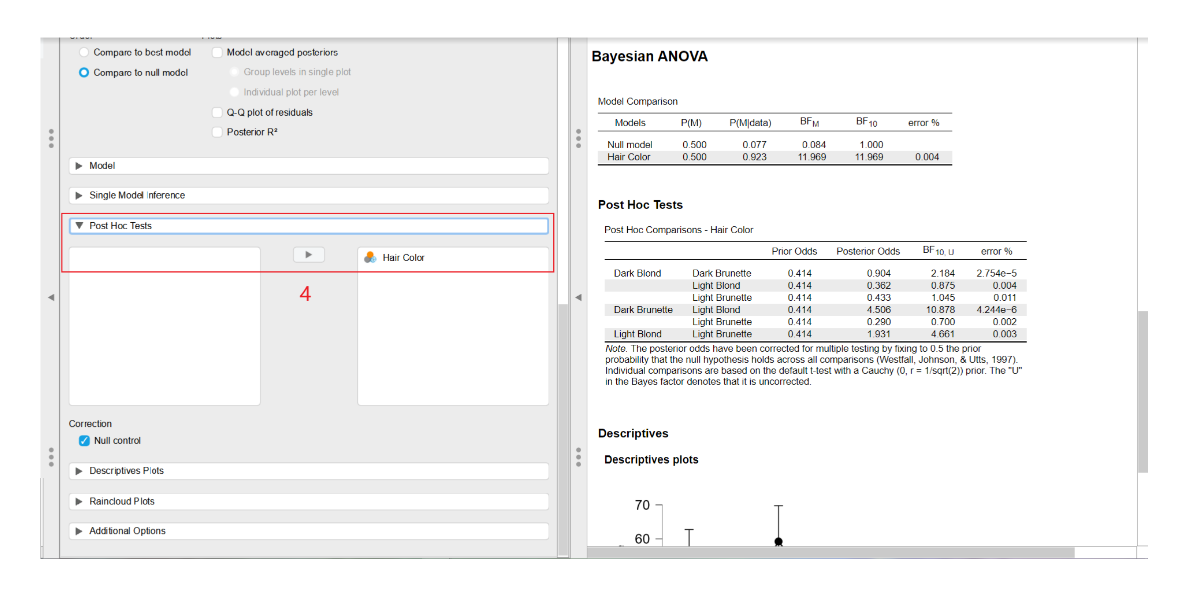


图 3 使用 JASP 进行事后检验的操作截屏

结果汇报。贝叶斯 ANOVA 的结果显示：参考先前研究者对 BF 数值大小所代表意义的划分(Jarosz & Wiley, 2014; Jeffreys, 1961; Wetzels et al., 2011)，贝叶斯因子 $BF_{incl} = 11.97$，说明在当前数据中 H_1 出现的可能性是 H_0 的 11.97 倍。这是较强的证据支持了 H_1，即不同发色的个体有着不同的疼痛容忍度。通过事后检验发现，当前数据中出现浅色金发个体(Light Blond)对疼痛的容忍度高于深色黑发个体(Dark Brunette)的可能性是二者无差异的 10.88 倍($BF_{10} = 10.88$，较强证据支持 H₁)；浅色金发个体(Light Blond)对疼痛的容忍度高于浅色黑发个体(Light Brunette; $BF_{10} = 4.66$，中等程度证据支持 H_1)；深色金发个体(Dark Blond)对疼痛的容忍度高于深色黑发个体($BF_{10} = 2.18$，较弱证据支持 H_1)；其余条件两两比较之间无证据支持存在差异或者不存在差异(BF_{10} 均小于 1.05)。

3.2 二因素方差分析

示例数据。使用心率(Heart Rate)数据作为示例，该数据来自 2×2 组间设计的实验。自变量是性别(Gender)和组别(Group)，因变量为运动六分钟后的心率(见图 4)。因此在本例中，H_0 为不同的性别和组别对心率的影响无显著差异，H_1 为性别和组别不仅存在主效应，且二者的交互作用也存在。

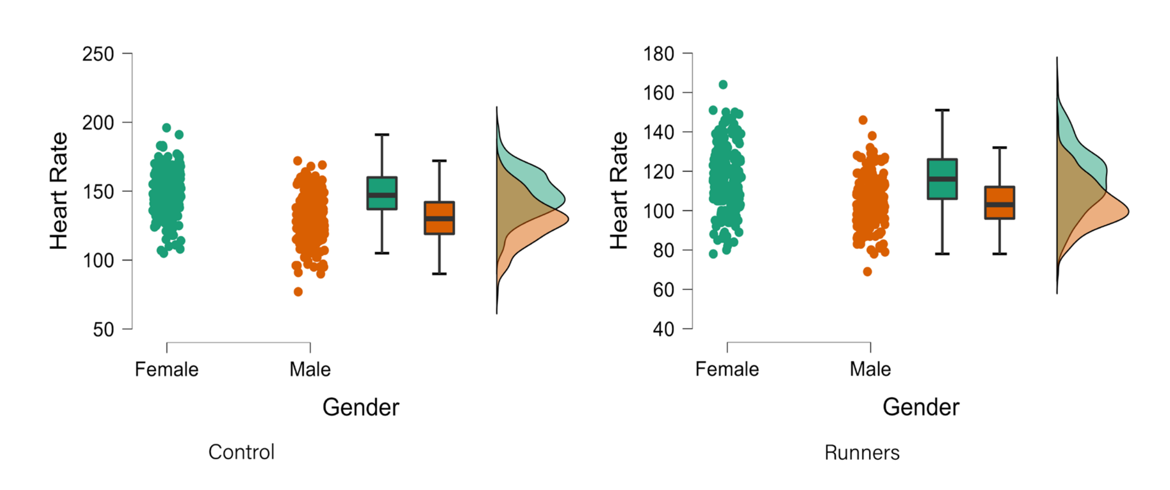
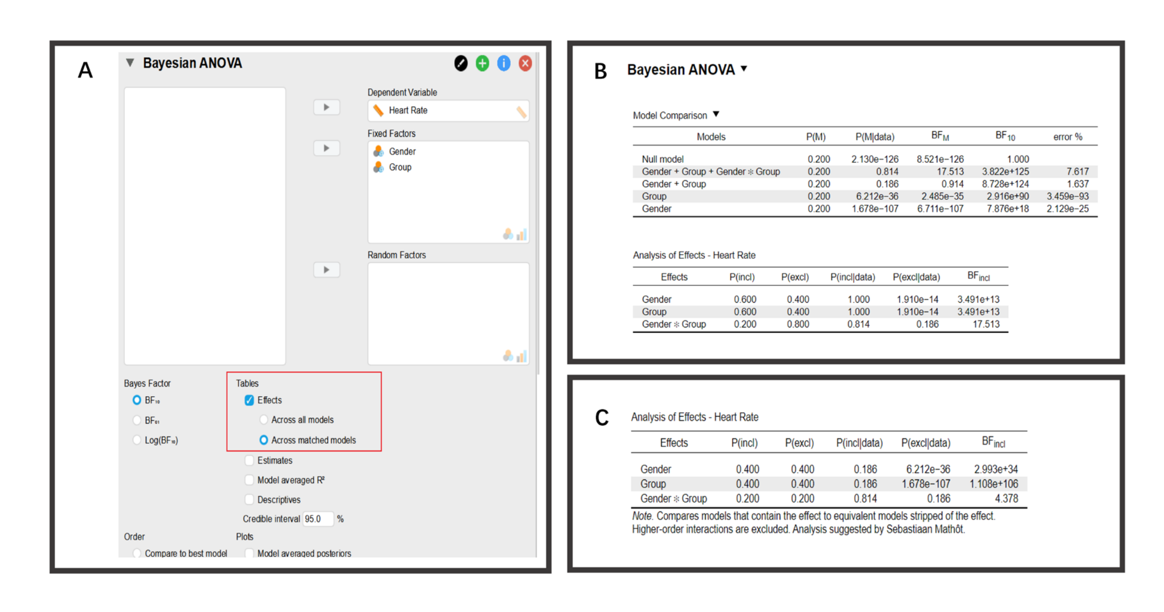


图 4 JASP 中示例数据心率的描述统计结果

操作步骤。首先在 JASP 中打开数据(Open - Data Library - ANOVA - Heart Rate)，然后在 ANOVA 面板中选择“Bayesian AVOVA”。之后需要进行以下步骤：第一，将心率作为因变量放入“Dependent Variable”中，将性别和组别作为自变量放入“Fixed Factors”中；第二，在“Bayes Factor”选项框中选择 BF_{10} 作为要输出的 BF 的形式；第三，在“order”选项框中选择 “compare to null model”；第四，在多因素实验设计中，如果需要计算BF_{incl}，就需要勾选“Effects”。并且选择“Across all models”和“Across matched models”方法所计算出的 BF_{incl} 是不同的。本文建议采取 Sebastiaan Mathôt 提倡的“Across matched models”方法计算的结果(见图 5A)；第五，由于贝叶斯 ANOVA 没有简单效应分析的模块，如果研究中需要进行进一步简单效应分析，可通过贝叶斯 t 检验进行简单效应分析。

具体而言，“Effects”的选择会影响 BF_{incl} 的计算。在进行 BF_{incl} 的计算前，首先要了解如何计算$P(incl)$、$P(excl)$、$P(incl|data)$和$P(excl|data)$。在 JASP 中，这四个值的计算有两种：包括所有模型的取向(across all models)和包括匹配模型的取向(across matched models)。两种取向的选择在 JASP 中的操作见图 5A。这两种取向之下的计算公式有所区别，具体计算过程见在线补充材料 E。



B Bayesian ANOVA

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.200	2.130e−126	8.521e−126	1.000	
Gender + Group + Gender ✻ Group	0.200	0.814	17.513	3.822e+125	7.617
Gender + Group	0.200	0.186	0.914	8.728e+124	1.637
Group	0.200	6.212e−36	2.485e−35	2.916e+90	3.459e−93
Gender	0.200	1.678e−107	6.711e−107	7.876e+18	2.129e−25

Analysis of Effects - Heart Rate

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BF_{incl}
Gender	0.600	0.400	1.000	1.910e−14	3.491e+13
Group	0.600	0.400	1.000	1.910e−14	3.491e+13
Gender ✻ Group	0.200	0.800	0.814	0.186	17.513

C Analysis of Effects - Heart Rate

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BF_{incl}
Gender	0.400	0.400	0.186	6.212e−36	2.993e+34
Group	0.400	0.400	0.186	1.678e−107	1.108e+106
Gender ✻ Group	0.200	0.200	0.814	0.186	4.378

Note. Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.

图 5 A 为选择“Effects”；B 为选择 Across all models 方法后的 BF_{incl} 输出结果；C 为选择 Across matched models 方法后的 BF_{incl} 输出结果

结果汇报。贝叶斯 ANOVA 的结果显示：参考先前研究者对 BF 数值大小所代表意义的划分(Jarosz & Wiley, 2014; Jeffreys, 1961; Wetzels et al., 2011)，当前数据存在极强的证据支持性别主效应(BF_{incl} = 2.99×10^{34})和组别主效应(BF_{incl} = 1.11×10^{106})的存在。女性运动六分钟后的心率(M = 131.99, SD = 22.72)高于男性(M = 116.99, SD = 19.84)，控制组运动六分钟后的心率(M = 139.00, SD = 18.95)高于跑步组(M = 109.98, SD = 15.53)。当前数据存在中等程度的证据支持性别和组别间交互作用的存在(BF_{incl} = 4.38)。

分别对控制组和跑步组进行贝叶斯独立样本 t 检验，结果显示：在控制组中，女性运动六分钟后的心率高于男性(BF_{10} = 5.56×10^{20})；在跑步组中，女性运动六分钟的心率高于男性(BF_{10} = 5.48×10^{12})。

3.3 单因素重复测量方差分析

示例数据。使用来自于单因素组内设计实验的 Bush Tucker Food 数据作为示例。自变

量是食物种类，因变量为吃下食物后感到反胃需要的时间(秒)。数据的组织形式采用宽数据(wide data)的形式，与 SPSS 中进行分析时的数据组织形式相同。

操作步骤。首先在 JASP 中打开数据(Open - Data Library - ANOVA -Bush Tucker Food)，然后在 ANOVA 面板中选择“Bayesian Repeated Measures AVOVA”。之后需要进行以下步骤(见图 6)：第一，设置新的重复测量因子，为每个水平进行命名；第二，在“Bayes Factor”选项框中选择 BF_{10} 作为要输出的 BF 的形式；第三，在“order”选项框中选择 “compare to null model”；第四，由于该示例数据中的自变量食物种类是一个四水平组内变量，因此需要事后检验来观测差异究竟出现在哪两个水平之间。

结果汇报。对结果的汇报可以参考先前对单因素方差分析的汇报形式。

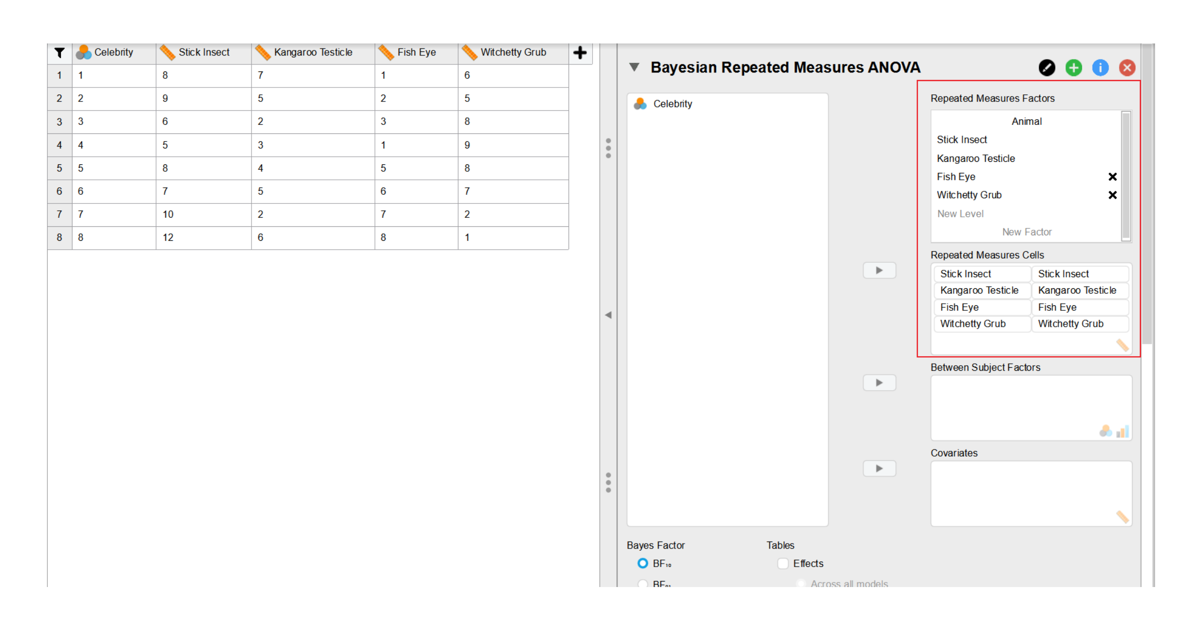


图 6 使用 JASP 进行单因素重复测量方差分析的操作截屏

3.4 二因素重复测量方差分析

3.4.1 实验设计为组内设计

示例数据。酒精态度(Alcohol Attitudes)采用的是 3×3 组内设计。自变量是图片效价(Imagery)和饮品种类(Drink)，各有三个水平，因变量为被试在观看图片后对饮品的态度评

分(见图 7)。

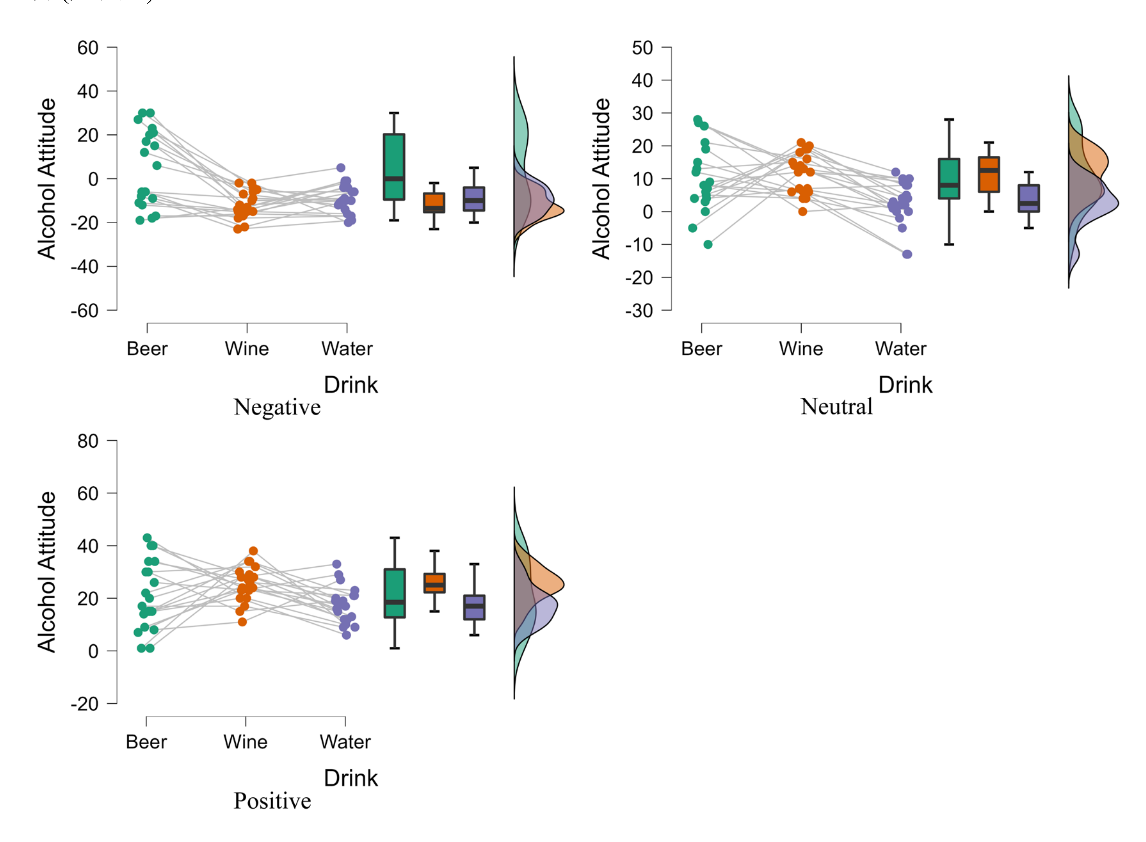


图 7 示例数据酒精态度的描述统计结果

操作步骤。首先在 JASP 中打开数据(Open - Data Library - ANOVA - Alcohol Attitudes)，然后在 ANOVA 面板中选择“Bayesian Repeated Measures AVOVA”。之后需要进行以下步骤(见图 8)：第一，设置两个重复测量因子并对不同因子的不同水平进行命名；第二，在“Bayes Factor”选项框中选择 BF_{10} 作为要输出的贝叶斯因子的形式；第三，在“order”选项框中选择“compare to null model”；第四，在“Tables”选项框中勾选“Across matched models”来计算 BF_{incl}；第五，由于贝叶斯 ANOVA 没有简单效应分析的模块，因此需要通过贝叶斯 t 检验进行简单效应分析。

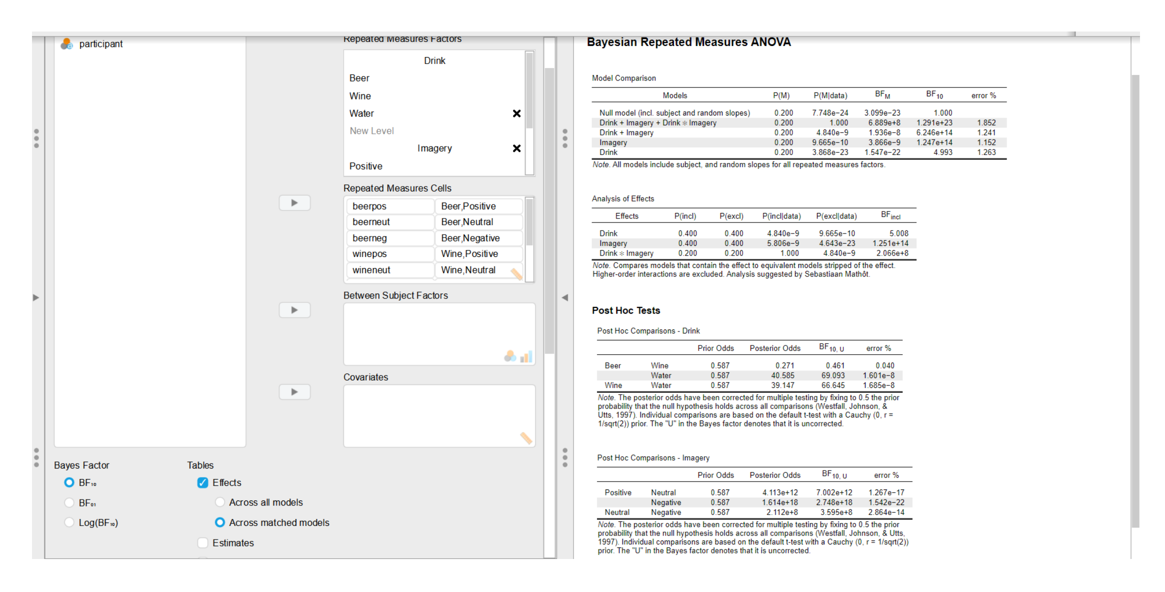


图 8 使用 JASP 进行二因素重复测量方差分析的操作截屏

结果汇报。贝叶斯 ANOVA 的结果显示：参考先前研究者对 BF 数值大小所代表意义的划分(Jarosz & Wiley, 2014; Jeffreys, 1961; Wetzels et al., 2011)，当前数据存在中等程度的证据支持饮品种类主效应($BF_{\text{incl}} = 5.01$)和极强的证据支持图片效价主效应($BF_{\text{incl}} = 1.25\times10^{14}$)的存在。事后检验发现，水的态度得分低于啤酒($BF_{10} = 69.09$)和红酒($BF_{10} = 66.65$)，啤酒和红酒的态度得分无差异($BF_{10} = 0.46$)；积极图片的态度得分高于消极图片($BF_{10} = 2.75\times10^{18}$)和中性图片($BF_{10} = 7.00\times10^{12}$)，中性图片的态度得分高于消极图片($BF_{10} = 3.60\times10^{8}$)。当前数据存在极强的证据支持图片效价和饮品种类间交互作用的存在($BF_{\text{incl}} = 2.07\times10^{8}$)。

3.4.2 实验设计为混合设计

示例数据。举重速度采用的是 2×3 混合设计。自变量是抓握类型(Grip)和负重(RM)，因变量为举重速度。该数据需在 https://jasp-stats.org/teaching-with-jasp/中进行下载。

操作步骤。首先在 JASP 中打开数据，然后在 ANOVA 面板中选择“Bayesian Repeated Measures AVOVA”。之后需要进行以下步骤(见图 9)：第一，设置重复测量因子并对因子的不同水平进行命名，并将抓握类型放入“Between Subject Factors”中；第二，在“Bayes Factor”

选项框中选择 BF_{10} 作为要输出的贝叶斯因子的形式；第三，在“order”选项框中选择“compare to null model”；第四，在“Tables”选项框中勾选“Across matched models”来计算 BF_{incl}；第五，由于贝叶斯 ANOVA 没有简单效应分析的模块，因此需要通过贝叶斯 T-test 进行简单效应分析。

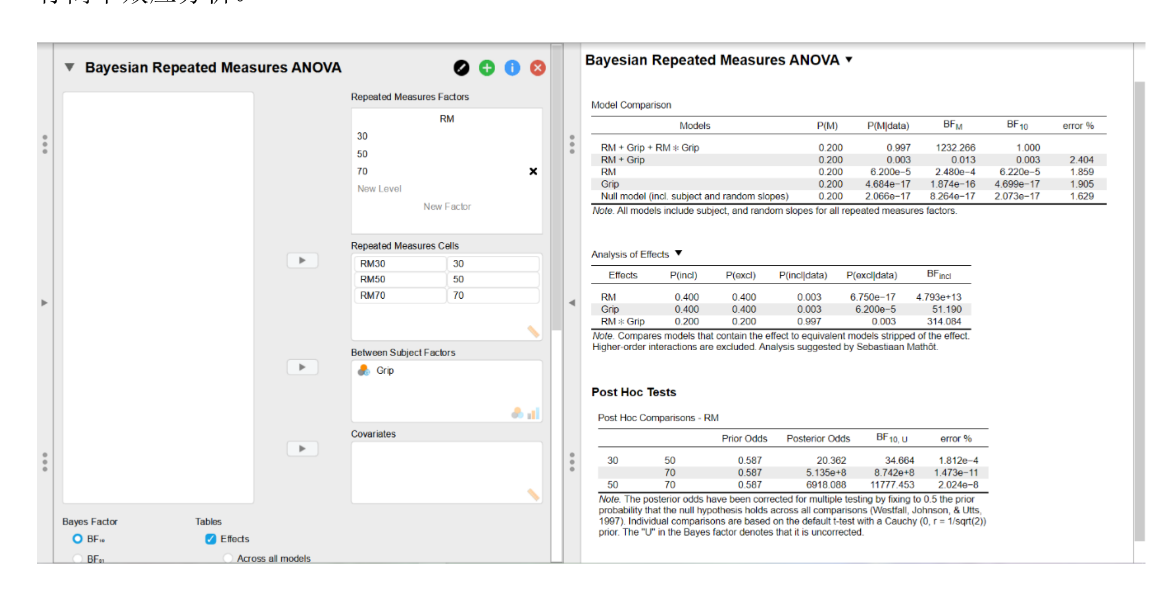


图 9 使用 JASP 进行二因素重复测量方差分析的操作截屏

结果汇报。贝叶斯方差分析的结果显示：参考先前研究者对 BF 数值大小所代表意义的划分(Jarosz & Wiley, 2014; Jeffreys, 1961; Wetzels et al., 2011)，当前数据存在非常强的证据支持抓握类型主效应的存在($BF_{\mathrm{incl}} = 51.19$)，传统抓握方式的举重速度快于反握。当前数据存在极强的证据支持负荷主效应的存在($BF_{\mathrm{incl}} = 4.79 \times 10^{13}$)的存在。事后检验发现，30%负荷的举重速度快于 50%负荷($BF_{10} = 34.66$)和 70%负荷($BF_{10} = 8.74 \times 10^{8}$)，50%负荷的举重速度快于 70%负荷($BF_{10} = 11777.45$)。当前数据存在极强的证据支持抓握方式和负荷间交互作用的存在($BF_{\mathrm{incl}} = 314.08$)。

4 讨论

本文介绍了贝叶斯 ANOVA 的基本原理，特别是使用线性模型并进行模型比较的思维

方式。随后结合实验心理学常用的五种实验设计，本文展示了如何在 JASP 中进行贝叶斯 ANOVA 及如何汇报并解释统计结果，为贝叶斯 ANOVA 的使用提供示例。

本文主要关注如何理解贝叶斯 ANOVA 以及如何在 JASP 中进行操作，因此未深入对以下内容进行讨论。第一，先验分布的参数设定。本文的所有示例均使用 JASP 的默认设定 (Prior: r scale fixed effects = 0.5, r scale random effects = 1, r scale covariates = 0.354; Model Prior: Uniform)，当研究者觉得有必要进行修改时，可在 Additional Options 中进行修改。由于修改这些先验本身即是对混合线性模型中先验的修改，研究者可使用 R 中的 brms 包来构建贝叶斯混合效应模型(潘晚坷 等, 2022)，该方式也更加灵活。第二，要对参数先验和模型先验进行区分，虽然本文在原理部分进行了介绍，但这仍然是容易混淆之处，需要研究者特别注意。第三，汇报 BF 时要汇报实际数值，而非简单的与特定阈值进行比较(Schmalz et al., 2021)。例如，不推荐仅汇报 $BF > 3$，而是汇报具体数值。如果需要更为全面的了解在 JASP 中进行贝叶斯统计的细节，可以参考 van Doorn 等(2021)的文章。

van den Bergh 等(2020)认为贝叶斯 ANOVA 还存以下两个需要注意的问题：第一，贝叶斯 ANOVA 与频率学派 NHST 存在同样的问题，即当模型被错误指定并且残差分布是非正态分布时，其结果可能会出现偏差。该问题可通过使用不指定残差分布的 ANOVA(例如，Kruskal-Wallis 检验)或者指定残差分布来解决(需在 Stan 或 JAGS 中进行)；第二，BF 的计算会受到模型内参数先验分布的影响。BF 实质上是两个模型边际似然的比值，先验分布的变化必然会导致 BF 的变化(Schad et al., 2022; Tendeiro & Kiers, 2019)。在复杂的实验设计中，研究者在了解并设置合适的参数先验分布是困难的。因此，本文推荐使用 JASP 默认的先验分布参数设定。

贝叶斯统计的优势足以令心理学等相关学科的研究者将注意从频率学派 ANOVA 转移到贝叶斯 ANOVA(胡传鹏 等, 2018; 李贵玉, 顾昕, 2021; 许岳培 等, 2022; 郑元瑞, 胡传鹏,

2023)。贝叶斯 ANOVA 可以起到补充和检验频率学派 ANOVA 结果的作用，从而为研究结果提供更为有力的支持(Hoijtink et al., 2019)。BF 通常会使研究者得出和 p 值一致的结果。当结果不一致时，建议研究者按以下流程对结果进行报告：1.详细地报告 BF 和 p 值的结果及各自所代表的含义；2.研究者在做出结论时要持有谨慎态度，避免对结果的过度解读。当这两种统计方法得出不一致的结论时，可能有多个原因，例如，当前研究的效应量不够稳定，或者当前的数据不满足 ANOVA 分析的前提预设。这提示研究者需要反思当前研究，包括样本量是否足够、在实验中对无关变量的控制情况等。同时，这也启发研究者要在收集数据前确定关键效应的量值。基于此，贝叶斯因子序列分析为研究者提供了一个新的分析视角。该方法要求研究者在实验数据收集开始前，就要根据研究设计或假设确定关键效应。在保证研究获得足够证据的前提下，设置停止收集数据的规则。在数据收集的过程中，研究者可以对数据进行持续地分析，直到 BF 和样本量达到阈值，就可以停止数据收集(详细步骤和实现教程可参考：郑元瑞, 胡传鹏, 2023)。

BF 会比 p 值更加保守并提供更直观的信息(Dong & Wedel, 2017)。更为重要的是，随着贝叶斯统计的应用以及相关软件和软件包的开发(例如，JASP、Stan、JAGS、BayesFactor、brms、bain、BANOVA 和 PyMC 等)，研究者能更加快速、方便地开展贝叶斯分析。为了贝叶斯分析的稳定性和可重复性，研究者也提出了一系列分析指南(Schad et al., 2022; van Doorn et al., 2021)。总之，贝叶斯统计的应用能够为解决心理学的可重复危机做出重要贡献。

参考文献

胡传鹏, 孔祥祯, Wagenmakers, E. J., Alexander, L. Y., 彭凯平. (2018). 贝叶斯因子及其在 JASP 中的实现. *心理科学进展*, *26*(6), 951-965.

李贵玉, 顾昕. (2021). 贝叶斯统计方法的应用与现状. *心理学探新*, *41*(5), 466-473.

孟迎芳, 董月晴, 陈荃. (2021). 概念内隐记忆中的注意促进效应. *心理学报*, *53*(5), 469-480.

潘晚坷, 温秀娟, 金海洋. (2022). 贝叶斯混合效应模型在心理学中的应用教程. CSTR:32003.36.ChinaXiv.202210.00098.V3

吴凡, 顾全, 施壮华, 高在峰, 沈模卫. (2018). 跳出传统假设检验方法的陷阱——贝叶斯因子在心理学研究领域的应用. *应用心理学*, *24*(3), 195-202.

许岳培, 陆春雷, 王珺, 宋琼雅, 贾彬彬, 胡传鹏. (2022). 评估零效应的三种统计方法. *应用心理学*, *28*(3), 369-384.

郑元瑞, 胡传鹏. (2023). 贝叶斯因子序列分析：实验设计中平衡信息与效率的新方法. *应用心理学*.

Brydges, C. R., & Gaeta, L. (2019). An introduction to calculating Bayes factors in JASP for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4523-4533.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, *90*(432), 1313-1321.

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*, *96*(453), 270-281.

Derks, K., de Swart, J., Wagenmakers, E. J., Wille, J., & Wetzels, R. (2021). JASP for audit: Bayesian tools for the auditing practice. *Journal of Open Source Software*, *6*(68), 2733.

Dong, C., & Wedel, M. (2017). BANOVA: An R package for hierarchical Bayesian ANOVA. *Journal of Statistical Software*, *8*(9), 1-46.

Faulkenberry, T. J., Ly, A., & Wagenmakers, E.-J. (2020). Bayesian inference in numerical cognition: A tutorial using JASP. *Journal of Numerical Cognition*, *6*(2), 231-259.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2-18.

Grünwald, P., de Heide, R., & Koolen, W. M. (2020). Safe testing. In *2020 Information Theory and Applications Workshop (ITA)* (pp. 1-54), San Diego, CA, USA.

Gelman, A., Hill, J., & Yajima, M. (2012) Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211.

Heck, D. W., & Bockting, F. (2021). Benefits of Bayesian model averaging for mixed-effects modeling. *Computational Brain & Behavior*, *6*, 35-49.

Heck, D. W., Boehm, U., Boing-Messing, F., Burkner, P. C., Derks, K., Dienes, Z., … Hoijtink, H. (2022). A review of applications of the Bayes factor in psychological research. *Psychological Methods*, *28*(3), 558-579.

Hendriksen, A., de Heide, R., & Grünwald, P. (2021). Optional stopping with Bayes factors: A categorization and extension of folklore results, with an application to invariant situations. *Bayesian Analysis*, *16*(3), 961-989.

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E. J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200-215.

Hoijtink, H., Mulder, J., Van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the

Bayes factor. *Psychological Methods*, *24*(5), 539-556.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *The Journal of Problem Solving*, *7*(1), Article 2.

JASP Team. (2022). *JASP (Version 0.16.3)* [Computer software]. Retrieved from https://jasp-stats.org/

Jeffreys, H. (1961). *Theory of probability (3rd Ed.)*. Oxford, UK: Oxford University Press.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178-206.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, *103*(481), 410-423.

Morey, R. D., & Rouder, J. N. (2022). *BayesFactor: Computation of Bayes factors for common designs* [R package version 0.9.12-4.4]. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356-374.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E. J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*(2), 304-321.

Schad, D. J., Nicenboim, B., Burkner, P. C., Betancourt, M., & Vasishth, S. (2022). Workflow

techniques for the robust use of bayes factors. *Psychological Methods*. Advance online publication.

Schmalz, X., Biurrun Manresa, J., & Zhang, L. (2021). What is a Bayes factor? *Psychological Methods*. Advance online publication.

Srinivasan, M. R., & Vijayaragunathan, R. (2021). Bayes factors for comparison of two-way ANOVA models. *Journal of Statistical Theory and Applications*, *19*(4), 540-546.

Tendeiro, J. N., & Kiers, H. A. L. (2019). A review of issues about null hypothesis Bayesian testing. *Psychological Methods*, *24*(6), 774-795.

van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., … Wagenmakers, E.-J. (2020). A tutorial on conducting and interpreting a bayesian ANOVA in JASP. *L'Année psychologique*, *120*(1), 73-96.

van den Bergh, D., Wagenmakers, E. J., & Aust, F. (2023). Bayesian repeated-measures analysis of variance: An updated methodology implemented in JASP. *Advances in Methods and Practices in Psychological Science*, *6*(2).

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., … Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813-826.

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, *60*(3), 158-189.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic*

Bulletin & Review, *25*(1), 58-76.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35-57.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426-432.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291-298.

The implementation of Bayesian ANOVA in JASP: A practical primer

WANG Yunhong[1], Don van den Bergh[2], Frederik Aust[2], Alexander Ly[3], Eric-Jan Wagenmakers[2], HU Chuanpeng[4]

([1]Department of Psychology, School of Education, Shanghai Normal University, Shanghai 200234, China)

([2]Department of Psychological Methods, University of Amsterdam, 1018 VZ Amsterdam, The Netherlands)

([3]Centrum Wiskunde & Informatica, 1090 GB Amsterdam, The Netherlands)

([4]School of Psychology, Nanjing Normal University, Nanjing, 210024, China)

Abstract: The application of Bayesian statistics to hypothesis testing - Bayes factors - is increasing in psychological science. Bayes factors quantify the evidence supporting the competing hypothesis or model, respectively, thereby making a judgment about which hypothesis or model is more supported by the data based on its value. The principles and applications of Bayes factor for ANOVA are, however, not available in China. We first present the theoretical foundation of Bayesian ANOVA and its calculation rules. It also shows how to perform Bayesian ANOVA and how to interpret and report the results of five common designs (one-factor between-group design, one-factor within-group design, two-factor between-group design, two-factor within-group design, and two-factor mixed design) using example data. Theoretically, Bayesian ANOVA is an effective alternative to conventional ANOVA as a powerful vehicle for statistical inferences.

Key words: Bayesian statistics; Bayes factors; ANOVA; JASP